

Clustering Penduduk Kurang Mampu Di Desa Mekar Baru Menggunakan Algoritma K-Means

Egy Andryan¹, Asrul Abdullah^{2*}, Putri Yuli Utami³

^{1,2}Teknik Informatika, Universitas Muhammadiyah Pontianak, Indonesia

¹gmadzaky@gmail.com

^{2*}e-mail: asrul.abdullah@unmuhpnk.ac.id

³e-mail: 181220053@unmuhpnk.ac.id

| Diterima | Direvisi | Disetujui |
|------------|------------|------------|
| 10-08-2024 | 14-08-2024 | 15-08-2024 |

Abstrak - Kemiskinan merupakan masalah yang telah lama ada dan belum berhasil diselesaikan oleh pemerintah. Meskipun berbagai upaya, seperti program bantuan sosial tunai (BLT), telah dilakukan, masih terdapat tantangan dalam pelaksanaannya. Salah satu kendala yang dihadapi pemerintah daerah di Desa Mekar Baru adalah ketidakmerataan dan ketidaktepatan sasaran dalam distribusi bantuan sosial. Algoritma *K-Means*, yang merupakan salah satu algoritma paling populer dan sederhana, digunakan untuk mengelompokkan data penduduk kurang mampu menjadi beberapa klaster. Berdasarkan hasil evaluasi, diperoleh empat klaster: klaster 1 (sangat tidak mampu) dengan 114 data (36,8% dari total data), klaster 0 (sangat mampu) dengan 89 data (28,8%), klaster 3 (mampu) dengan 83 data (26,8%), dan klaster 2 (sangat mampu) dengan 23 data (7,4%). Klaster tersebut diurutkan dari yang memiliki jumlah terbesar hingga terkecil. Penelitian ini bertujuan untuk mengembangkan sebuah aplikasi yang dapat mengidentifikasi klaster dari dataset yang dianalisis menggunakan metode K-Means.

Kata Kunci: *Maching Learning, klastering, K-means, Penduduk Kurang mampu*

Abstract - Poverty is a longstanding issue that the government has yet to resolve. Despite various efforts, such as the cash transfer program (BLT), challenges persist in its implementation. One of the difficulties faced by local governments in Desa Mekar Baru is the uneven and inaccurate distribution of social assistance. The *K-Means* algorithm, known for its simplicity and popularity, is used to cluster data on low-income residents into several groups. Based on the evaluation results, four clusters were identified: Cluster 1 (very poor) with 114 data points (36.8% of the total), Cluster 0 (very affluent) with 89 data points (28.8%), Cluster 3 (affluent) with 83 data points (26.8%), and Cluster 2 (very affluent) with 23 data points (7.4%). These clusters are ranked from the largest to the smallest. The goal of this study is to develop an application that can identify the clusters obtained from the dataset using the *K-Means* method.

Keywords: *Analytic Hierarchy Process, Decision Support System, Product*

PENDAHULUAN

Kemiskinan adalah salah satu masalah yang telah lama dihadapi oleh pemerintah, namun belum sepenuhnya terselesaikan. Berbagai upaya telah dilakukan, salah satunya melalui program Bantuan Sosial Tunai (BLT), yang bertujuan untuk mengurangi beban ekonomi masyarakat miskin. Namun, meskipun bantuan ini disalurkan, masih terdapat banyak tantangan dalam proses distribusinya. Salah satu permasalahan yang sering terjadi adalah penyaluran bantuan yang tidak merata dan tidak tepat sasaran. Hal ini terutama disebabkan oleh pendataan yang tidak akurat dan sistem distribusi yang masih manual. Berdasarkan wawancara dengan Ketua RT Desa Mekar

Baru, Bapak Ahmad, diketahui bahwa pendataan di desa ini masih dilakukan secara manual menggunakan *Microsoft Excel*, yang mengakibatkan lambatnya proses penentuan penerima bantuan (Suhartini & Yuliani, 2021).

Untuk mengatasi masalah ini, dibutuhkan sebuah sistem yang lebih efisien dan akurat. Sistem berbasis web diharapkan dapat membantu dalam proses penentuan penerima bantuan sosial dengan lebih cepat dan tepat. Dengan adanya sistem ini, diharapkan bantuan dapat disalurkan kepada mereka yang benar-benar membutuhkan, sehingga masalah ketidakmerataan dan ketidaktepatan sasaran dapat diminimalisir (Arianto, 2019)



Dalam konteks ini, data mining dapat digunakan untuk mengklasifikasikan penduduk berdasarkan tingkat kemampuan ekonominya. Salah satu metode data mining yang efektif untuk tugas ini adalah *K-Means Clustering*. Metode ini digunakan untuk mengelompokkan data penduduk ke dalam beberapa kluster berdasarkan kemiripan data mereka. Dengan metode ini, Kantor Desa Mekar Baru dapat mengidentifikasi kelompok-kelompok penduduk yang paling membutuhkan bantuan, sehingga bantuan sosial dapat disalurkan dengan lebih tepat sasaran (Filki, 2022).

Metode *K-Means Clustering* dipilih karena memiliki tingkat akurasi yang baik dan dapat menangani data dalam jumlah besar dengan efisien. Algoritma K-Means adalah bagian dari kelompok *Unsupervised Learning* yang digunakan untuk membagi data ke dalam beberapa kluster berdasarkan karakteristik tertentu. Dalam penelitian ini, penulis mengaplikasikan algoritma *K-Means* pada data penduduk kurang mampu di Desa Mekar Baru. Data yang digunakan mencakup 1.432 penduduk dari tahun 2022 yang diperoleh dari Kantor Kepala Desa (Suhartini & Yuliani, 2021).

Penelitian ini bertujuan untuk mengembangkan sebuah aplikasi web yang dapat mengklasifikasikan penduduk kurang mampu berdasarkan metode *K-Means*. Dengan adanya aplikasi ini, diharapkan dapat membantu Kantor Desa Mekar Baru dalam menyalurkan bantuan sosial dengan lebih tepat dan efektif, sehingga masalah kemiskinan dapat ditangani dengan lebih baik.

1. Data mining

Data Mining adalah suatu siklus yang memanfaatkan setidaknya satu prosedur pembelajaran PC untuk memeriksa dan memisahkan informasi secara alami atau serangkaian siklus untuk menghilangkan nilai tambah dari kombinasi informasi sebagai informasi yang sebelumnya tidak jelas secara fisik. Lantaran itu Data Mining sebenarnya mempunyai akar yang panjang berdasarkan bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik dan *database*. *Data mining* adalah sebuah proses untuk menentukan korelasi pada pola, dan tren baru yang bermakna dengan memilah-milah data dalam jumlah sangat besar yang disimpan di dalam repositori, menggunakan teknologi pengenalan pola serta teknik statistik dan matematika. *Data mining* adalah proses menemukan sebuah pola dan pengetahuan menarik dari data dalam jumlah yang sangat besar (Nabila et al., 2021)

2. Algoritma K-Means

Proses membagi sampel yang sama menjadi beberapa kelompok atau kluster yang sesuai disebut dengan *clustering*, yang merupakan sub bab dari data mining. Menemukan struktur dalam data yang sebelumnya tidak diketahui adalah tujuan dari clustering. Ada banyak perhitungan *Bunching* yang dapat diakses, seperti *K-Means*. Algoritma yang paling

sederhana dan paling banyak digunakan, *K-Means*, mengatur data ke dalam cluster yang telah ditentukan (Rafi Nahjan et al., 2023)

3. Metode elbow

Kelompok pembelajaran tanpa pengawasan mencakup algoritma *Clustering K-Means*, yang menggunakan sistem partisi untuk membagi data menjadi beberapa kelompok. Pembelajaran mandiri merupakan perhitungan penggalian informasi untuk mencari desain dari seluruh faktor (SKS), faktor (*ascribes*) yang tujuan/nilai/kelasnya tidak jelas (tidak ada). Perhitungan *K-Means* merupakan prosedur pengelompokan berbasis jarak yang memartisi informasi menjadi beberapa kelompok dan perhitungan ini hanya berhubungan dengan bilangan atau sifat numerik (Nabila et al., 2021)

Tentukan jumlah kluster menggunakan metode *elbow* untuk menemukan titik optimal. Pilih pusat kluster secara acak, tetapi dalam penelitian ini, pusat kluster ditentukan dari urutan tertinggi hingga terendah. Hitung jarak antara data dan pusat kluster menggunakan *Euclidean Distance*:

$$d(x_j, c_j) = \sum_{j=1}^n (x_j - c_j)^2 \dots \dots \dots (1)$$

Keterangan :

d = jarak

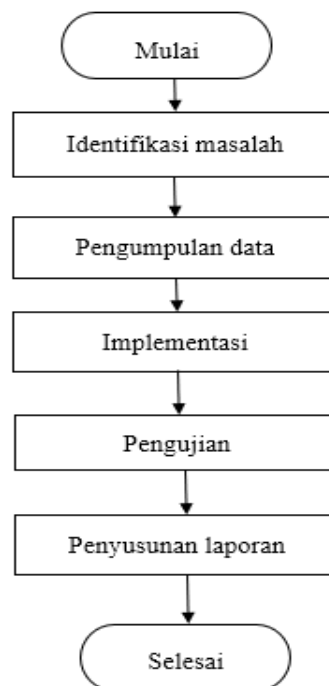
xj = Data ke-jjj

cj = centroid ke-j

Tempatkan data pada kluster terdekat, lalu hitung ulang pusat kluster. Ulangi Data Tidak berubah,

METODE PENELITIAN

Berikut ini adalah tahapan-tahapan yang akan peneliti gambarkan melalui alur metode penelitian pada Gambar 1 yang di bawah ini:



Sumber : Hasil Penelitian, 2023

Gambar 1. Metodologi Penelitian

a. Identifikasi Masalah

Identifikasi masalah dan merumuskan masalah pada penelitian untuk bertujuan mengatasi masalah yang di alami kantor Desa Mekar baru agar pemberian bantuan sosial kepada masyarakat kurang mampu dibagikan sangat tepat.

b. Pengumpulan Data

Pengumpulan data dalam penelitian ini mencakup data penduduk kurang mampu di Desa Mekar Baru. Proses pengumpulan data dilakukan melalui tiga tahapan: pertama, observasi langsung untuk memahami masalah yang ada; kedua, wawancara dengan pihak RT, yaitu Bapak Ahmad, untuk mendapatkan data primer mengenai penduduk kurang mampu melalui survei dan pengamatan langsung; ketiga, studi pustaka untuk memperkuat penelitian dengan mengacu pada buku dan jurnal yang relevan, khususnya terkait metode *Clustering K-Means* untuk menentukan jumlah kluster optimal.

c. Implementasi dan Pengujian

Proses Data Mining: Setelah *dataset* diperoleh, langkah awal yang dilakukan adalah proses data *mining* menggunakan metode *Knowledge Discovery in Database (KDD)*, di mana *preprocessing* dilakukan untuk mempersiapkan *dataset* sebelum dilatih oleh model guna menentukan jumlah kluster.

Perancangan Web: Pada tahap ini, penulis mendesain UI dan mengembangkan aplikasi web menggunakan bahasa pemrograman *Python*.

Clustering Algoritma *K-Means*: Algoritma *K-Means*, yang termasuk dalam kelompok *Unsupervised learning*, digunakan untuk mengelompokkan data ke dalam beberapa kelompok berdasarkan sistem partisi. *Unsupervised learning* adalah metode data *mining* yang mencari pola dari semua variabel tanpa menentukan target atau label.

Pengujian: Tahap ini melibatkan pengujian kluster menggunakan metode *cohesion* dan *separation*. *Cohesion* diukur dengan menghitung semua objek dalam sebuah kluster, sementara *separation* diukur dengan menghitung jarak rata-rata setiap objek dalam kluster dengan kluster terdekat, menggunakan rumus *Euclidean distance*.

d. Pengujian

Pengujian yang saat ini menggabungkan metode pemisahan dan kohesi adalah pengujian *cluster*. Jumlah objek dalam sebuah *cluster* dihitung untuk menentukan kohesi, dan jarak rata-rata antara masing-masing objek dalam sebuah *cluster* dan *cluster* terdekat digunakan untuk menentukan pemisahan. Jarak antar informasi ditentukan dengan menggunakan persamaan jarak *Euclidean*.

e. Penyusunan laporan

Laporan akhir penelitian ditulis dengan menggunakan kerangka kerja yang mencakup pendahuluan, landasan teori, metodologi penelitian, analisis, implementasi dan

pengujian sistem, penutup, serta lampiran hasil penelitian. Berdasarkan tabel ini untuk menjelaskan jenis kriteria penduduk kurang mampu pada tabel 3.1 yang di bawah ini:

Tabel 1. Tabel Kriteria penduduk kurang mampu

| NO | KRITERIA | KETERANGAN |
|----|---------------------------|--|
| 1 | Jenis kelamin | Jenis kelamin pada kepala keluarga. |
| 2 | Umur | Menentukan umur |
| 3 | Wni | Apakah asli Indonesia atau tidak |
| 4 | Jenis pekerjaan | Apakah memiliki pekerjaan yang tetap |
| 5 | Kepemilikan rumah | Memiliki tempat tinggal sendiri atau sewa |
| 6 | Jenis lantai | Memiliki jenis lantai yang ditinggali |
| 7 | Jenis atap | Memiliki jenis atap yang ditinggali |
| 8 | Jenis dinding | Memiliki jenis dinding yang ditinggali |
| 9 | Jenis bahan bakar memasak | Memiliki jenis bahan bakar memasak yang layak |
| 10 | Sumber air | Memiliki sumber air yang layak untuk digunakan sehari-hari |
| 11 | Jumlah kendaraan | Memiliki jumlah kendaraan yang cukup |
| 12 | Penghasilan per bulan | Penghasilan yang diperoleh per bulan |
| 13 | Pendidikan | Apakah pendidikan tamat atau tidak |
| 14 | ASN | Apakah menjadi pegawai negeri atau tidak |
| 15 | KTP | Apakah penduduk pindahan atau penduduk asli |

Sumber : Penelitian 2023

HASIL DAN PEMBAHASAN

Bab ini berisi uraian tentang hasil, analisis dan pengujian aplikasi *clustering* penduduk kurang mampu di Desa Mekar Baru menggunakan Algoritma *K-means*. Pengujian ini akan mencari hasil cluster terbaik menggunakan *silhouette score*.

a. Hasil *Preprocessing*

Pada tahapan ini akan dilakukan *preprocessing* merupakan proses untuk mempersiapkan data sebelum dilatih oleh model.

Berikutnya yaitu proses Label Encoding.

Label *Encoding* merupakan proses mengonversi data-data kategorial menjadi numerik. Berikut Gambar 5.1 proses label *encoding*.

Sumber: Penelitian 2023

Gambar 2. Label *Encoding*

Berdasarkan Gambar 2 di atas maka proses label *encoding* untuk mengubah data kategorial ke numerik agar dapat mudah dipahami oleh model.

Berikutnya yaitu proses *missing value*. *Missing value* merupakan kondisi dimana ada satu atau sejumlah data tidak tersedia atau hilang dalam sebuah set data. Berikut ini adalah informasi terkait *missing value* dari semua fitur pada Gambar 3 di bawah ini.

```

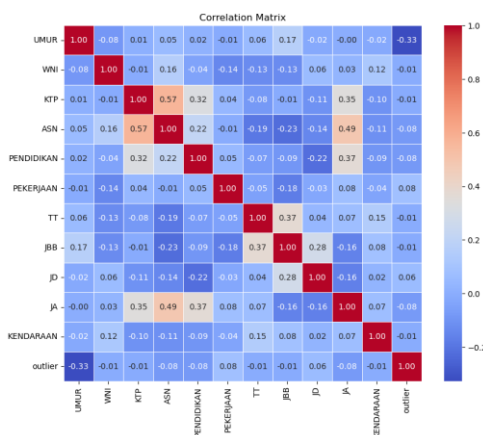
---
UMUR          0
WNI           0
KTP           0
ASN           0
PENDIDIKAN   0
PEKERJAAN    0
TT           0
JBB          0
JD           0
JA           0
KENDARAAN    0
dtype: int64
    
```

Sumber: Penelitian 2023

Gambar 3. Informasi *Missing Value*

Berdasarkan Gambar 5.2 diatas adalah kategori mana saja yang terdapat *missing value*. Setelah melakukan analisis terdapat *missing value*, dapat dilihat bahwa ada data yang hilang yaitu jenis kelamin, jumlah pendapatan, jenis lantai.

Setelah itu ada kerangka koneksi. Dibandingkan dengan penelitian sebelumnya, matriks korelasi digunakan untuk meningkatkan akurasi. Matriks korelasi digunakan untuk mencari kumpulan data untuk atribut-atribut yang berkorelasi dan dapat dihilangkan. Visualisasi *correlation matrix* dapat dilihat pada gambar 4 di bawah ini.

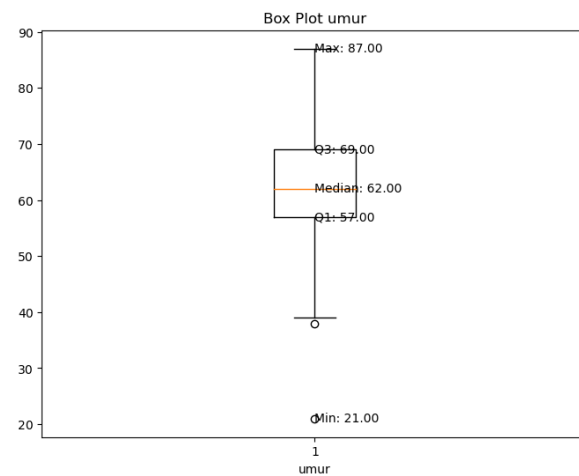


Sumber: Penelitian 2023

Gambar 4. *Correlation Matrix*

Pada Gambar 4 diatas dapat dilihat nilai korelasi antar atribut dalam skala warna. Angkat di dalam kotak-kotak adalah nilai korelasi. Maka nilai korelasi maka nilai tertinggi atribut KTP dan ASN dengan nilai 0,57. Nilai korelasi mendekati 1 menunjukkan korelasi positif, mendekati=1 menunjukkan korelasi negatif dan nilai mendekati 0 korelasi lemah.

Terakhir *Ouliter*, *Outlier* atau bisa juga dikenal dengan anomali ini merupakan sebuah data atau observasi yang menyimpang secara ekstrem dari rata-rata sekumpulan data yang ada. Penyimpangan ini bisa berupa angkanya terlalu tinggi atau terlalu rendah. Cara untuk melihat data *outlier* bisa dengan *box plot* seperti Gambar 5 dibawah ini.



Sumber: Penelitian 2023

Gambar 5. Informasi *Missing Value*

Pada Gambar 5 di atas merupakan hasil dari penanganan menggunakan nilai median pada data umur tanpa ubah data apa pun agar tidak merusak atau mempengaruhi data lainnya.

b. Implementasi

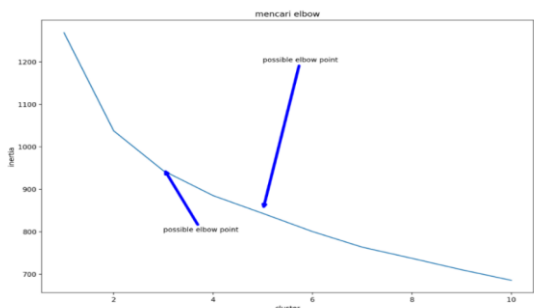
Untuk mencari nilai yang ideal, dilakukan strategi perampingan dengan menggunakan teknik siku untuk mencari jumlah tandan terbaik. Berikutnya adalah efek samping dari estimasi SSE sehubungan dengan pengujian kualitas k=2 hingga k=10.

Tabel 2. Tabel Kriteria penduduk kurang mampu

| | |
|-----|----------|
| K2 | 63.65000 |
| K3 | 55.61702 |
| K4 | 68.38461 |
| K5 | 44.33333 |
| K6 | 46.52380 |
| K7 | 21.00000 |
| K8 | 51.00000 |
| K9 | 59.59016 |
| K10 | 66.00000 |

Sumber: Penelitian 2023

Dari perhitungan SSE metode *elbow* pada Tabel 5. Di atas didapatkan nilai SSE yang tertinggi terdapat pada nilai k=4. Sehingga jumlah *cluster* terbaik yang dibentuk sejumlah 4 *cluster*.



Sumber: Penelitian 2023

Gambar 6. Informasi Missing Value

Berdasarkan Gambar 6 di atas dapat dilihat bahwa pada k=4 mengalami penurunan jika dibandingkan jumlah k lainnya. Jumlah cluster selain k=4 mengalami penurunan yang stabil. Maka dapat ditentukan jumlah cluster optimal yang terbentuk 4 cluster.

Pilih pusat Cluster secara acak. Pemilihan pusat klaster dapat dilakukan dengan berbagai cara, yang paling sering dilakukan adalah secara acak. Dengan menentukan 4 cluster maka dipilih centroid awal atau pusat cluster yaitu data ke 9 sebagai pusat cluster 1, data ke 14 sebagai pusat cluster 2, data ke 41 sebagai pusat cluster 3 dan data ke 114 sebagai pusat cluster 4 yang akan digunakan untuk perhitungan iterasi ke-1.

Tabel 2. Menentukan pusat centroid awal(bagian kiri)

| Da ta | centr oid | um ur | w ni | k t p | a s n | pen didi kan | peker jaan |
|----------|--------------|----------|---------|-------------|-------------|--------------------|---------------|
| 9 | C1 | 68 | 2 | 1 | 2 | 1 | 1 |
| 14 | C2 | 48 | 1 | 2 | 2 | 2 | 2 |
| 41 | C3 | 60 | 2 | 1 | 1 | 1 | 2 |
| 11 4 | C4 | 68 | 1 | 1 | 1 | 2 | 1 |

Sumber : Penelitian 2023

Tabel 3. Menentukan pusat centroid awal(bagian kanan)

| PP | JL | TT | SA | BBM | JD | JA | KDRAN |
|----|----|----|----|-----|----|----|-------|
| 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 2 | 2 | 1 | 2 | 2 | 2 | 1 |
| 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 |
| 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 |

Sumber : Penelitian 2023

Estimasi pemisahan dari informasi 1 ke tempat grup sesuai persamaan pada 4.6 di atas. Berikutnya adalah ilustrasi perhitungan antar informasi

1(69,1,2,2,1,1,2,1,2,2,2,2,2) ke centroid awal.

Menghitung data ke centroid:

Melakukan perhitungan jarak terhadap jumlah penduduk kurang mampu dengan data pertama ke titik centroid cluster pertama.

$$\text{Data (1.1)} = (69-68)^2+(1-2)^2+(2-1)^2+(2-2)^2+(1-1)^2+(1-1)^2+(2-2)^2+(1-1)^2+(2-1)^2+(2-1)^2+(2-1)^2+(2-1)^2+(2-1)^2+(2-1)^2 = 21.1187121$$

Dan dihitung terus sampai dengan Data 309.

Melakukan perhitungan jarak terhadap jumlah penduduk kurang mampu dengan data pertama ke centroid cluster kedua.

$$\text{Data (1.2)} = (69-48)^2+(1-1)^2+(2-2)^2+(2-2)^2+(1-2)^2+(1-2)^2+(2-2)^2+(1-2)^2+(2-2)^2+(2-1)^2+(2-2)^2+(2-2)^2+(2-2)^2+(2-2)^2+(2-1)^2 = 9.433981132$$

Dan dihitung terus sampai dengan Data 309.

Perhitungan jarak jumlah penduduk kurang mampu dengan data pertama ke centroid cluster ketiga.

$$\text{Data (1.3)} = (69-60)^2+(1-2)^2+(2-1)^2+(2-1)^2+(1-2)^2+(1-1)^2+(2-2)^2+(1-2)^2+(2-2)^2+(2-2)^2+(2-2)^2+(2-1)^2+(2-1)^2+(2-2)^2 = 2.449489743$$

Dan dihitung terus sampai dengan Data 309.

Perhitungan jarak jumlah penduduk kurang mampu dengan data pertama ke centroid ke empat.

$$\text{Data (1.4)} = (69-60)^2+(1-2)^2+(2-1)^2+(2-1)^2+(1-2)^2+(1-1)^2+(2-2)^2+(1-2)^2+(2-2)^2+(2-2)^2+(2-2)^2+(2-1)^2+(2-1)^2+(2-2)^2 = 2.4494897$$

Tabel berikut akan dibuat setelah seluruh data penduduk kurang mampu dari data 1 sampai dengan data 309 dihitung untuk setiap pusat klaster pada iterasi pertama dengan menggunakan rumus Eucliden Distance, di antaranya:

Tabel 4. Hasil Perhitungan Jarak Semua Data Ke Pusat Cluster Pada Iterasi 1

| No | C1 | C2 | C3 | C4 | Jarak Terde kat | Clu ster |
|----|----------------|-----------------|-----------------|-----------------|-----------------------|-------------|
| 1 | 21.11 87121 | 9.4339811 32 | 2.44948 9743 | 2.44948 97 | 2.4494 897 | 4 |
| 2 | 26.13 42687 | 14.212670 4 | 6.55743 8524 | 6.55743 852 | 6.5574 385 | 4 |
| 3 | 27.05 54985 | 15.198684 15 | 7.34846 9228 | 7.34846 923 | 7.3484 692 | 4 |
| 4 | 20.22 37484 | 8.4852813 74 | 2.64575 1311 | 2.64575 1311 | 2.6457 513 | 4 |
| 5 | 24.12 46762 | 12.124355 65 | 4.47213 5955 | 4.47213 5955 | 4.4721 360 | 3 |
| 6 | 21.23 67606 | 9.0553851 38 | 2.64575 1311 | 2.64575 1311 | 2.6457 513 | 3 |
| 7 | 27.14 77439 | 15.033296 38 | 7.41619 8487 | 7.41619 8487 | 7.4161 985 | 4 |
| 8 | 20.09 97512 | 8.5440037 45 | 2.44948 9743 | 2.44948 9743 | 2.4494 897 | 3 |
| 9 | 24.10 39416 | 12.165525 06 | 5 0000 | 5.00000 0000 | 5.0000 000 | 4 |
| 10 | 17.14 64282 | 5.7445626 47 | 3.74165 7387 | 3.74165 7387 | 3.7416 574 | 3 |
| 11 | 21.14 23745 | 9.3808315 2 | 1 0000 | 1.00000 0000 | 1.0000 000 | 4 |

| | | | | | | |
|-------|----------------|-----------------|-----------------|------------------|---------------|-------|
| 12 | 19.18 33261 | 7.5498344 35 | 3.16227 766 | 3.16227 7660 | 3.1622 777 | 3 |
| | | | | | | |
| 300 | 7.615 77311 | 5.5677643 63 | 13.1909 0596 | 13.1909 05958 | 5.5677 644 | 2 |
| 301 | 12.20 65556 | 2.8284271 25 | 8.30662 3863 | 8.30662 3863 | 2.8284 271 | 1 |
| 302 | 10.39 23048 | 3.3166247 9 | 10.4880 8848 | 10.4880 88482 | 3.3166 248 | 1 |
| 303 | 10.19 8039 | 3.8729833 46 | 10.2956 3014 | 10.2956 30141 | 3.8729 833 | 2 |
| 304 | 15.23 15462 | 4.1231056 26 | 5.47722 5575 | 5.47722 5575 | 4.1231 056 | 3 |
| 305 | 13.11 4877 | 3 4877 | 7.48331 4774 | 7.48331 4774 | 3.0000 000 | 2 |
| 306 | 13.26 64992 | 2.6457513 11 | 7.48331 4774 | 7.48331 4774 | 2.6457 513 | 2 |
| 307 | 8.426 14977 | 4.8989794 86 | 12.2882 0573 | 12.2882 05727 | 4.8989 795 | 1 |
| 308 | 6.557 43852 | 6.4807406 98 | 14.3178 2106 | 14.3178 21063 | 6.4807 407 | 2 |
| 309 | 15.19 86842 | 4 86842 | 5.38516 4807 | 5.38516 4807 | 4.0000 000 | 2 |

Sumber : Penelitian 2023

Setelah masing-masing informasi dimasukkan ke dalam kelompok terdekat dan hasilnya diketahui pada penekanan pertama, maka tempat kelompok yang lain akan dihitung ulang berdasarkan berapa banyak informasi yang ada. Untuk mendapatkan komunitas grup lain untuk C2, C3 dan C4, lakukan sesuai perhitungan C1 di atas. Hasil estimasi perolehan tempat grup baru pada penekanan kedua adalah sebagai berikut:

Tabel 5. Pusat Cluster Baru Untuk Iterasi Ke-2(bagian kiri)

| cent roid | umur | Wni | k t p | asn | pendi dikan | peker jaan | PP |
|-----------|--------------|--------------|-------|--------------|--------------|------------|--------------|
| C1 | 4.88 3027 | 3.22 0339 | 0 | 0.15 2484 | 0.073 847 | 0 | 0.00 1292 |
| C2 | 5.04 1582 | 1 | 0 | 0 | 0 | 0 | 0 |
| C3 | 3.73 415 | 2.05 1282 | 0 | 0 | 0 | 0 | 0 |
| C4 | 4.33 432 | 2.44 23 | 0 | 0 | 0 | 0 | 0 |

Sumber : Penelitian 2023

Tabel 6. Pusat Cluster Baru Untuk Iterasi Ke-2(bagian kanan)

| JL | T T | SA | BBM | JD | JA | KDR AN |
|--------------|-----|-----------------|-------------|--------------|--------------|--------------|
| 0.000 626 | 0 | 1.0951 1E-05 | 5.3E- 06 | 0 | 9.28 E-08 | 0.035 783 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 4.222 34 | 3.322 442 | 2.444 22 | 0 |

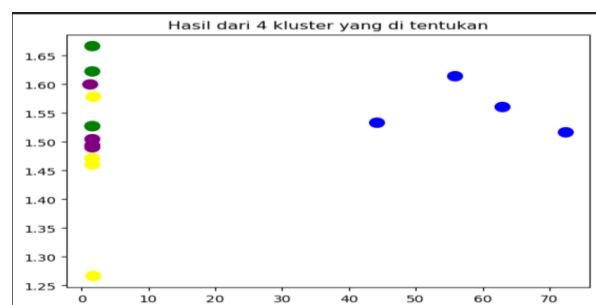
Sumber : Penelitian 2023

Pada Tabel 5 dan 6 di atas akan dilakukan estimasi ke depannya dari setiap informasi populasi,

khususnya tabel 4 untuk tempat kelompok baru pada tabel 5 dan 6. Perhitungan untuk menentukan tempat kelompok baru dilakukan seperti pada model di atas. sampai tandannya tidak berubah lagi maka estimasi dihentikan.

Penentuan Hasil Klaster yang Stabil: Pada tahap ini, perhitungan dihentikan pada iterasi kedua karena data penduduk tidak mengalami perubahan pada iterasi tersebut.

Hasil Pengelompokan Algoritma K-Means: Halaman ini menampilkan hasil pengelompokan menggunakan algoritma K-Means, yang memvisualisasikan klaster yang dihasilkan. Dataset yang digunakan diperoleh dari Kantor Desa Mekar Baru, dan hasilnya dapat dilihat pada Gambar 7.



Sumber: Penelitian 2023

Gambar 7. Hasil Pengelompokan

Penelitian melakukan perhitungan dengan menggunakan persamaan untuk menghitung jarak antar data pada K-means menggunakan rumus Euclidian Distance. Pada halaman ini ditampilkan hasil jumlah cluster yang menghasilkan 4 cluster. Untuk lebih jelasnya dapat dilihat. Gambar 8 di bawah ini.



Sumber: Penelitian 2023

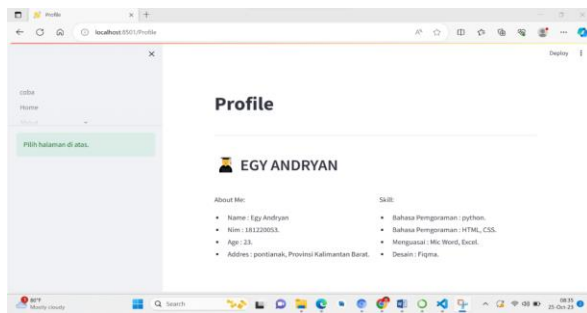
Gambar 8. Hasil Pengelompokan

Berdasarkan Gambar 7 di atas maka menunjukkan anggota penduduk pada setiap cluster dengan nilai k=4. Dimana dari jumlah 309 data penduduk menjadi anggota pada cluster 1 sejumlah 114 anggota penduduk, cluster 0 sejumlah 89 anggota penduduk, cluster 3 sejumlah 83 anggota penduduk dan cluster 2 anggota sejumlah 23 penduduk.

Pada tahapan berikutnya yaitu integrasi cara menyimpan dan menggunakan model yang disimpan, Menyimpan pemodelan ipynb yang dibangun ke dalam file csv menggunakan Library Python dengan code di bawah ini.

```
df = pd.read_csv('dpm.csv', index_col=0)
df.head()
```

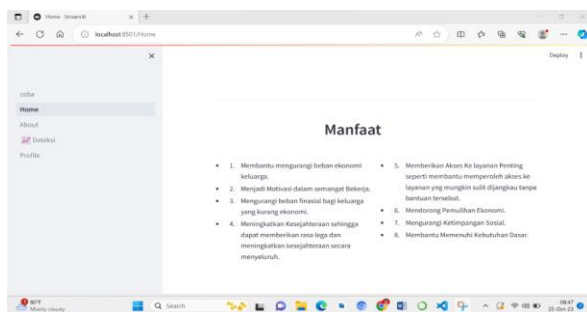
Hasil Halaman Profile Pada halaman ini ditampilkan biodata penulis dari pembangunan website yaitu berisi tentang Nama, Nim, Age, Address. Serta menampilkan sebuah kelebihan. Untuk lebih jelasnya Halaman Profile dapat dilihat Gambar 9 dibawah ini.



Sumber: Penelitian 2023

Gambar 9. Halaman Profile

Berikutnya Hasil halaman *home* Pada halaman ini ditampilkan hasil dari pembangunan *website* yaitu berisi beberapa informasi tentang manfaat penerimaan bantuan, halaman *home* dapat dilihat Gambar 10 di bawah ini.



Sumber: Penelitian 2023

Gambar 10. Halaman Home

c. Evaluasi

Pada tahap pengujian dengan menggunakan *silhouette coefficient* bahwa untuk mencari nilai terbaik dalam setiap *cluster*. Pada Gambar 11 dibawah ini.



Sumber: Penelitian 2023

Gambar 11. Hasil Evaluasi *Silhouette Coefficient*

Berdasarkan Gambar 11, hasil pengujian menunjukkan bahwa kolom '*Cluster*' memiliki empat nilai unik, yaitu 1, 0, 3, dan 2, dengan proporsi kemunculan yang berbeda-beda. *Cluster* 1, yang mewakili kategori sangat tidak mampu, muncul paling banyak dengan 36.8% dari total data. *Cluster* 0, yang mewakili kategori sangat mampu, muncul sebanyak 28.8%. *Cluster* 3, yang mewakili kategori mampu, muncul sebanyak 26.8%, sedangkan *Cluster* 2, yang mewakili kategori sangat mampu, muncul paling sedikit dengan 7.4%. Hasil ini menggambarkan distribusi dan persentase masing-masing kategori dalam kolom '*Cluster*'.

KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa penulis berhasil membangun sebuah aplikasi untuk mengidentifikasi *cluster* dalam *dataset* menggunakan metode K-Means. Pengujian dengan *silhouette score* pada *dataset* penduduk kurang

mampu menghasilkan empat *cluster* dengan distribusi sebagai berikut: *Cluster* 1 (sangat tidak mampu) mencakup 114 data (36.8%), *Cluster* 0 (sangat mampu) mencakup 89 data (28.8%), *Cluster* 3 (mampu) mencakup 83 data (26.8%), dan *Cluster* 2 (sangat mampu) mencakup 23 data (7.4%). Nilai *cluster* diurutkan dari yang terbesar hingga terkecil. Sebagai saran, penelitian selanjutnya disarankan untuk menggunakan *dataset* yang lebih besar untuk representasi *klasterisasi* yang lebih akurat, mempertimbangkan metode *klasterisasi* lain seperti DBSCAN atau *hierarchical clustering*, dan menambahkan metode validasi tambahan seperti *cross-validation* untuk memastikan validitas hasil pengujian.

REFERENSI

- Suhartini, S., & Yuliani, R. (2021). Penerapan Data Mining untuk Mengcluster Data Penduduk Miskin Menggunakan Algoritma K-Means di Dusun Bagik Endep Sukumulia Timur. *Infotek: Jurnal Informatika dan Teknologi*, 4(1), 39–50. <https://doi.org/10.29408/jit.v4i1.2986>
- Arianto, J. (2019). Penerapan Data Mining Untuk Pengelompokan Penduduk Kurang Mampu Desa Sambirejo Timur Dengan Algoritma K-Medoids (Studi Kasus Kantor Kepala Desa Sambirejo Timur). *KOMIK: Konferensi Nasional Teknologi Informasi dan Komputer*, 3(1), 569–573. <https://doi.org/10.30865/komik.v3i1.1660>
- Jurnal, M., Informasi, S., Misi, J., Manajemen, J., & Informasi, S. (2021). Volume 4, No 1, Januari 2021 ISSN: 2614-1701 (Cetak) – 2614-3739 (Online). *MISI: Jurnal Manajemen Informatika & Sistem Informasi*, 4(1).
- Filki, Y. (2022). Algoritma K-Means Clustering dalam Memprediksi Penerima Bantuan Langsung Tunai (BLT) Dana Desa. *Jurnal Informasi Ekonomi Bisnis*, 4, 166–171. <https://doi.org/10.37034/infkeb.v4i4.166>
- Bangsa, S. T., & Utara, S. (2021). Implementasi Data Mining Dalam Mengelompokkan Jumlah Penduduk Miskin Berdasarkan Provinsi Menggunakan Algoritma K-Means. *Jurnal Ilmiah Teknologi Informasi*, 2(2), 125–132.
- Sunia, D., Kurniabudi, & Jusia, P. A. (2019). Penerapan Data Mining Untuk Clustering Data Penduduk Miskin Menggunakan Algoritma K-Means. *Jurnal Ilmiah Mahasiswa Teknik Informasi*, 1(2), 121–134.
- Triyana, M., Juita, R., & Suhendra, C. D. (2022). Penerapan Metode K-Means dalam Pengelompokan Data Penduduk Tidak Mampu di Distrik Oransbari. *Jurnal Teknologi dan Sistem Informasi*, 7(3).
- Febriansyah, F., & Muntari, S. (2023). Penerapan Algoritma K-Means untuk Klasterisasi Penduduk Miskin pada Kota Pagar Alam. *Jurnal Teknologi Informasi*, 8(1), 66–77.
- Susanto, J. (2023). Klasterisasi Data Penerima

- Bantuan Langsung Tunai Menggunakan Algoritma K-Means. *Jurnal Komputer dan Informatika*, 10(2), 461–470.
<https://doi.org/10.30865/jurikom.v10i2.5807>
- Nabila, Z., Isnain, A. R., & Abidin, Z. (2021). Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means. *Jurnal Teknologi dan Sistem Informasi*, 2(2), 100. Retrieved from <http://jim.teknokrat.ac.id/index.php/JTSI>
- Nahjan, M. R., Heryana, N., & Voutama, A. (2023). Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell. *Jurnal Mahasiswa Teknik Informasi*, 7(1), 101–104.
<https://doi.org/10.36040/jati.v7i1.6094>
- Serikat, A., & Kunci, K. (2023). K-Means Clustering Dalam Pengelompokan Kabupaten Dan Kota Di Jawa Timur. *Seminar Nasional Hasil Riset dan Pengabdian (SNHRP)*, 863–870. Retrieved from <http://karyailmiah.unipasby.ac.id/wp-content/uploads/2019/04/K-Means-Artikel.pdf>
- Paembonan, S., & Abduh, H. (2021). Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat. *PENA Teknik: Jurnal Ilmiah Ilmu-Ilmu Teknik*, 6(2), 48.
https://doi.org/10.51557/pt_jiit.v6i2.659
- Hastomo, W., Aini, N., Karno, A. S. B., & Rere, L. M. (2022). Metode Pembelajaran Mesin untuk Memprediksi Emisi Manure Management. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, 11(2), 131–139.
<https://doi.org/10.22146/jnteti.v11i2.2586>